Joint Filtering Scheme for Nonstationary Noise Reduction

Jensen, Jesper Rindom; Benesty, Jacob; Christensen, Mads Græsbøll; Jensen, Søren Holdt

Published in:
Proceedings of the European Signal Processing Conference

Publication date:
2012

Document Version
Accepted author manuscript, peer reviewed version

JOINT FILTERING SCHEME FOR NONSTATIONARY NOISE REDUCTION

Jesper Rindom Jensen[1]*, Jacob Benesty*[3]*, Mads Græsbøll Christensen*[2]*, and Søren Holdt Jensen*[1]

[1]Aalborg University
Dept. of Electronic Systems
{jrj,shj}@es.aau.dk

[2]Aalborg University
Dept. of Architecture, Design & Media Technology
mgc@create.aau.dk

[3]University of Quebec
INRS-EMT
benesty@emt.inrs.ca

ABSTRACT

In many state-of-the-art filtering methods for speech enhancement, an estimate of the noise statistics is required. However, the noise statistics are difficult to estimate when speech is present and, consequently, nonstationary noise has a detrimental impact on the performance of most noise reduction filters. We propose a joint filtering scheme for speech enhancement which supports the estimation of the noise statistics even during voice activity. First, we use a pitch driven linearly constrained minimum variance (LCMV) filter to estimate the noise statistics. A Wiener filter is then designed based on the estimated noise statistics, and it is applied for the noise reduction of the speech. In experiments involving real signals, we show that the proposed filtering scheme outperforms other existing speech enhancement methods in terms of perceptual evaluation of speech quality (PESQ) scores in different nonstationary noise scenarios.

Index Terms— Speech enhancement, time-domain filtering, Wiener filter, LCMV filter, orthogonal decomposition, harmonic decomposition, pitch, nonstationary noise.

1. INTRODUCTION

Speech is frequently encountered in numerous signal processing applications such as telecommunications, teleconferencing, hearing-aids, and human-machine interfaces. The speech picked up by a microphone can be very noisy. Unfortunately, the noise will degrade the speech quality and intelligibility which, eventually, has a detrimental impact on speech applications. It is therefore highly relevant to develop methods for reducing the noise. In this paper, we consider filtering methods for noise reduction of single-channel speech recordings. Several such methods have been developed in the past decades. For an overview of such methods, we refer to [1] and the references therein. Many existing noise reduction filtering methods assume that the noise signal is directly available since they rely on the noise statistics. This is, of course, not the case in practice, so the noise statistics could, for example, be estimated when there is no voice activity. Some alternative methods based on, e.g., harmonic tunneling [2] and minimum statistics [3] have been proposed for estimating the noise statistics during speech presence.

In this paper, we propose a novel joint filtering scheme for nonstationary noise reduction of noisy quasi-periodic signals such as voiced speech. It is well-known that speech can be both voiced and unvoiced, so the proposed filtering scheme has to be combined with voiced/unvoiced speech detection (see, e.g., [4, 5]) when applied to speech enhancement. In the proposed scheme, we utilize two recently proposed filters, namely the orthogonal decomposition based Wiener (ODW) filter and the harmonic decomposition based linearly constrained minimum variance (HDLCMV) filter [6, 7]. Followingly, the proposed filtering scheme is described. First, we use the HDLCMV filter to obtain a rough estimate of the desired signal. In the HDLCMV filter, it is assumed that the desired signal is quasi-periodic and thereby has a harmonic structure which is a reasonable assumption for the voiced parts of speech signals. Therefore, the HDLCMV filter is designed using the pitch, the number of harmonics, and the statistics of the observed signal, i.e., this filter does not rely on noise statistics. Pitch and model order estimation is not considered in this paper, but there exists a multitude of methods for this (see, e.g., [7] and the references therein). From the rough estimate of the desired signal, we obtain an estimate of the noise signal. That is, using this approach, we can easily estimate the noise statistics even when speech is present. The estimated noise statistics are used to design the ODW filter which, finally, performs the noise reduction of the observed speech signal. Besides proposing the joint filtering scheme, we also provide a few important closed-form performance measure expressions for the filters under the assumption that the desired signal is quasi-periodic.

The remainder of the paper is organized as follows. In Section 2, we introduce the signal model used in the paper, the problem of designing noise reduction filters, and the orthogonal and harmonic decompositions. Based on this, we derive the optimal ODW and HDLCMV filters in Section 3. We then propose a joint filtering scheme for noise reduction and evaluate its performance in Section 4. Finally, in Section 5, we conclude on the paper.

2. SIGNAL MODEL

In this paper, we consider nonstationary noise reduction of single-channel speech recordings using filtering. The noise reduction problem is to extract a zero-mean desired signal, $x(n)$, from a mixture signal

$$y(n) = x(n) + v(n) \,, \tag{1}$$

where $v(n)$ is a zero-mean noise source, and n is the discrete time index. The noise source is assumed to be uncorrelated

with the desired signal. Moreover, in some parts of the paper, we assume that the desired signal is quasi-periodic which is indeed a reasonable assumption for, e.g., voiced speech. When the desired signal is quasi-periodic, we can rewrite the signal model in (1) as

$$y(n) = \sum_{l=1}^{L} \left(a_l e^{jl\omega_0 n} + a_l^* e^{-jl\omega_0 n} \right) + v(n) \,, \qquad (2)$$

where ω_0 is the fundamental frequency (aka the pitch), L is the model order, $a_l = \frac{A_l}{2} e^{j\phi_l}$ is the complex amplitude of the lth harmonic, A_l is the real amplitude of the lth harmonic, ϕ_l is the random phase of the lth harmonic, and $(\cdot)^*$ denotes the complex conjugation. Many real-life signals, however, have some degree of inharmonicity. The problem of inharmonicity is not considered in this paper, yet several methods dealing with it exist (see, e.g., [7] and the references therein).

When designing optimal filters for noise reduction, we need several consecutive samples of the observed signal, $y(n)$. Therefore, we use the vector signal model given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) \,, \qquad (3)$$

where

$$\mathbf{y}(n) = \begin{bmatrix} y(n) & y(n-1) & \cdots & y(n-M+1) \end{bmatrix}^T \,, \quad (4)$$

with $(\cdot)^T$ denoting the transpose of a vector or matrix, M is the number of samples, and the definitions of $\mathbf{x}(n)$ and $\mathbf{v}(n)$ follow that of $\mathbf{y}(n)$. We know by assumption that the desired signal and the noise are uncorrelated. Therefore, we can obtain the following simple expression for the covariance matrix,

$$\mathbf{R}_{\mathbf{y}} = \mathrm{E}[\mathbf{y}(n)\mathbf{y}^T(n)] = \mathbf{R}_{\mathbf{x}} + \mathbf{R}_{\mathbf{v}} \,, \qquad (5)$$

of the observed signal where $\mathrm{E}[\cdot]$ is the mathematical expectation operator, $\mathbf{R}_{\mathbf{x}} = \mathrm{E}[\mathbf{x}(n)\mathbf{x}^T(n)]$ is the covariance matrix of $\mathbf{x}(n)$, and $\mathbf{R}_{\mathbf{v}} = \mathrm{E}[\mathbf{v}(n)\mathbf{v}^T(n)]$ is the covariance matrix of $\mathbf{v}(n)$. When the desired signal is quasi-periodic, we can model the covariance matrix of $\mathbf{x}(n)$ as

$$\mathbf{R}_{\mathbf{x}} \approx \mathbf{Z}\mathbf{P}\mathbf{Z}^H \,, \qquad (6)$$

where $(\cdot)^H$ denotes the complex conjugate transpose of a matrix or vector, and

$$\mathbf{P} = \mathrm{diag}\left(\begin{bmatrix} |a_1|^2 & |a_1^*|^2 & \cdots & |a_L|^2 & |a_L^*|^2 \end{bmatrix} \right) \,, \quad (7)$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}(\omega_0) & \mathbf{z}^*(\omega_0) & \cdots & \mathbf{z}(L\omega_0) & \mathbf{z}^*(L\omega_0) \end{bmatrix}, \quad (8)$$

$$\mathbf{z}(l\omega_0) = \begin{bmatrix} 1 & e^{-jl\omega_0} & \cdots & e^{-jl\omega_0(M-1)} \end{bmatrix}^T \,, \quad (9)$$

with $\mathrm{diag}(\cdot)$ denoting the construction of a diagonal matrix from a vector.

The goal in noise reduction filtering methods is to design a filter which extracts one or more samples of the desired signal $x(n)$ from $\mathbf{y}(n)$. That is, the filter should attenuate the noise $v(n)$ as much as possible while not distorting the desired signal too much. In this paper, we focus on optimal filtering methods for extraction of a single sample of $x(n)$. A filtering operation which estimates $x(n)$ from $\mathbf{y}(n)$ can be written as

$$\hat{x}(n) = \sum_{m=0}^{M-1} h_m y(n-m) = \mathbf{h}^T \mathbf{y}(n) \,, \qquad (10)$$

where $\mathbf{h} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{M-1} \end{bmatrix}^T$, and $\hat{x}(n)$ is an estimate of $x(n)$. The main difference between optimal filtering methods for noise reduction is how the desired signal is decomposed. In this paper, we consider the orthogonal and harmonic decompositions [6, 7].

In the orthogonal decomposition, the signal vector $\mathbf{x}(n)$ is decomposed into two parts being proportional and orthogonal to $x(n)$, respectively. That is, using this decomposition, $\mathbf{x}(n)$ can also be written as

$$\mathbf{x}(n) = x(n)\boldsymbol{\rho}_{\mathbf{x}x} + \mathbf{x}_{\mathrm{i}}(n) = \mathbf{x}_{\mathrm{d}}(n) + \mathbf{x}_{\mathrm{i}}(n) \,, \qquad (11)$$

where

$$\boldsymbol{\rho}_{\mathbf{x}x} = \frac{\mathrm{E}[\mathbf{x}(n)x(n)]}{\mathrm{E}[x^2(n)]} \qquad (12)$$

is the normalized correlation vector between $\mathbf{x}(n)$ and $x(n)$. If we insert (11) into (10), we get

$$\begin{aligned} \hat{x}_{\mathrm{OD}}(n) &= \mathbf{h}^T[\mathbf{x}_{\mathrm{d}}(n) + \mathbf{x}_{\mathrm{i}}(n) + \mathbf{v}(n)] \\ &= x_{\mathrm{fd}}(n) + x_{\mathrm{ri}}(n) + v_{\mathrm{rn}}(n) \,, \end{aligned} \qquad (13)$$

where $x_{\mathrm{fd}}(n) = \mathbf{h}^T\mathbf{x}_{\mathrm{d}}(n)$ is the filtered desired signal, $x_{\mathrm{ri}}(n) = \mathbf{h}^T\mathbf{x}_{\mathrm{i}}(n)$ is the residual interference, and $v_{\mathrm{rn}}(n) = \mathbf{h}^T\mathbf{v}(n)$ is the residual noise. Using the orthogonal decomposition, we can define the following error signal

$$e_{\mathrm{OD}}(n) = x(n) - [x_{\mathrm{fd}}(n) + x_{\mathrm{ri}}(n) + v_{\mathrm{rn}}(n)] \,. \qquad (14)$$

Optimal noise reduction filters based on the orthogonal decomposition can then, for example, be derived by minimizing $e(n)$ or parts of $e(n)$ subject to some constraints. Most commonly, the error is minimized in the mean-square error (MSE) sense. Clearly, this design procedure ensures that the aforementioned design goals are fulfilled.

In the harmonic decomposition approach, it is assumed that the desired signal is quasi-periodic which makes it useful for signals produced by voiced speech and musical instruments [7, 8]. Due to this assumption, the signal vector, $\mathbf{x}(n)$, can be written as

$$\mathbf{x}(n) = \mathbf{Z}\mathbf{a}(n) = \mathbf{x}'_{\mathrm{d}}(n) \,, \qquad (15)$$

where

$$\begin{aligned} \mathbf{a}(n) = \begin{bmatrix} a_1 e^{j\omega_0 n} & a_1^* e^{-j\omega_0 n} & \cdots \\ & & a_L e^{jL\omega_0 n} \;\; a_L^* e^{-jL\omega_0 n} \end{bmatrix}^T . \end{aligned} \qquad (16)$$

From the above expression, we can see that there is no interference in this decomposition as opposed to in the orthogonal decomposition. This is because all information in $\mathbf{x}(n)$ can be used to describe the desired signal when we know the signal model. We can obtain an estimate of $x(n)$ using a harmonic decomposition filter by inserting (15) into (10). This yields

$$\hat{x}_{\text{HD}}(n) = \mathbf{h}^T[\mathbf{x}'_{\text{d}}(n) + \mathbf{v}(n)] \, . \tag{17}$$

We define the following error function for the harmonic decomposition approach to filter design

$$e_{\text{HD}}(n) = x(n) - [x'_{\text{fd}}(n) + v_{\text{rn}}(n)] \ , \tag{18}$$

where $x'_{\text{fd}}(n) = \mathbf{h}^T\mathbf{x}'_{\text{d}}(n)$. We can then design a harmonic decomposition based filter for noise reduction by minimizing the effects of $e_{\text{HD}}(n)$ or parts of $e_{\text{HD}}(n)$ perhaps subject to some constraints (e.g., to avoid undesired distortion).

3. OPTIMAL FILTERS

In this section, we derive the ODW filter and the HDLCMV filter. Furthermore, we provide expressions for the filters and some of their performance measures; the performance measure expressions are closed-form when the desired signal is periodic.

3.1. Orthogonal Decomposition Wiener

The ODW filter is found by minimizing $\text{E}\{|e_{\text{OD}}(n)|^2\}$ with respect to the unknown filter response. This yields

$$\mathbf{h}_{\text{W}} = \sigma_x^2\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x} \ , \tag{19}$$

where σ_x^2 is the variance of $x(n)$. When the desired signal is periodic, we can also write the normalized correlation vector, $\boldsymbol{\rho}_{\mathbf{x}x}$, as

$$\boldsymbol{\rho}_{\mathbf{x}x} = \frac{\mathbf{R}_{\mathbf{x}}\mathbf{i}}{\mathbf{i}^T\mathbf{R}_{\mathbf{x}}\mathbf{i}} = \frac{\mathbf{Z}\mathbf{P}\mathbf{1}}{\sigma_x^2} \ , \tag{20}$$

where $\mathbf{1} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T$ and $\mathbf{i}$ is the first column of the $M \times M$ identity matrix. That is, for periodic signals, the OD Wiener filter is given by

$$\mathbf{h}_{\text{W}} = \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1} \, . \tag{21}$$

The output signal-to-noise ratio (oSNR) of an orthogonal decomposition based filter is defined as the ratio between the variance of the filtered desired signal and the sum of the variances of the residual interference and noise [6]. It can be shown that the ODW filter achieves the maximum output SNR [6]. The output SNR of the ODW filter for periodic signals therefore equals

$$\text{oSNR}^{\text{OD}}(\mathbf{h}_{\text{W}}) = \frac{\mathbf{1}^T\mathbf{P}\mathbf{Z}^H\mathbf{R}_{\text{in}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1}}{\sigma_x^2} \ , \tag{22}$$

where $\mathbf{R}_{\text{in}} = \mathbf{R}_{\mathbf{x}_{\text{i}}} + \mathbf{R}_{\mathbf{v}}$ and $\mathbf{R}_{\mathbf{x}_{\text{i}}}$ is the covariance matrix of $\mathbf{x}_{\text{i}}(n)$. The harmonic distortion measure is useful when the desired signal is periodic. This measure is defined as the sum of the absolute differences between the harmonics before and after filtering. The harmonic distortion of the ODW filter can be show to be

$$\xi_{\text{hd}}(\mathbf{h}_{\text{W}}) = 2\sum_{l=1}^{L} P_l \left| 1 - |\mathbf{1}^T\mathbf{P}\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{z}(l\omega_0)|^2 \right| \ . \tag{23}$$

3.2. Harmonic Decomposition LCMV

This filter is designed for noise reduction of periodic signals. The HDLCMV filter is designed such that the variance of the residual noise is minimized under the constraint that the harmonics of the desired signal are not distorted. This design can also be written as the following optimization problem

$$\min_{\mathbf{h}} \mathbf{h}^T\mathbf{R}_{\mathbf{v}}\mathbf{h} \quad \text{s.t.} \quad \mathbf{Z}^H\mathbf{h} = \mathbf{1} \, . \tag{24}$$

The solution to the quadratic optimization problem above is well-known and given by

$$\mathbf{h}_{\text{HDLCMV}} = \mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}_{\mathbf{v}}\mathbf{Z}\right)^{-1}\mathbf{1} \tag{25}$$

$$= \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}\mathbf{Z}\right)^{-1}\mathbf{1} \, . \tag{26}$$

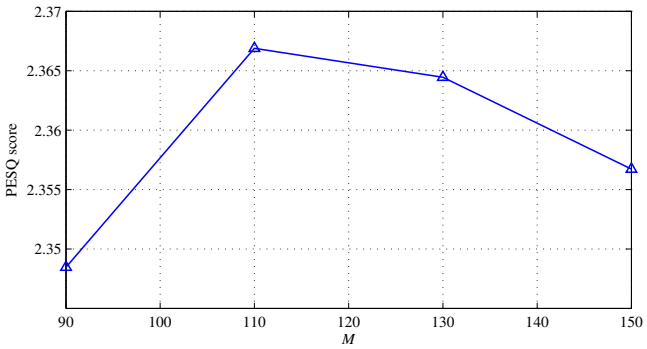
The step from (25) to (26) can be shown by using the matrix inversion lemma. From this expression, we can see that if we know the pitch ω_0 and the number of harmonics L then we only need the statistics of the observed signal $\mathbf{R}_{\mathbf{y}}$ to design the HDLCMV filter. Note that these parameters can be estimated using the very same HDLCMV filtering method [7]. In the ODW filter, we need to know either the statistics of the desired signal $\boldsymbol{\rho}_{\mathbf{x}x}$ or of the noise $\boldsymbol{\rho}_{\mathbf{v}v}$. When the filter order M becomes large and the desired signal is indeed periodic, it can be shown that the ODW and HDLCMV filters become identical. In the harmonic decomposition, there is is no interference term. The output SNR of a harmonic decomposition based filter is therefore simply defined as the ratio between the variances of the filtered desired signal and the residual noise. Therefore, the output SNR of the HDLCMV filter is given by

$$\text{oSNR}^{\text{HD}}(\mathbf{h}_{\text{HDLCMV}}) = \frac{\sigma_x^2}{\mathbf{1}^T(\mathbf{Z}^H\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Z})^{-1}\mathbf{1}} \ , \tag{27}$$

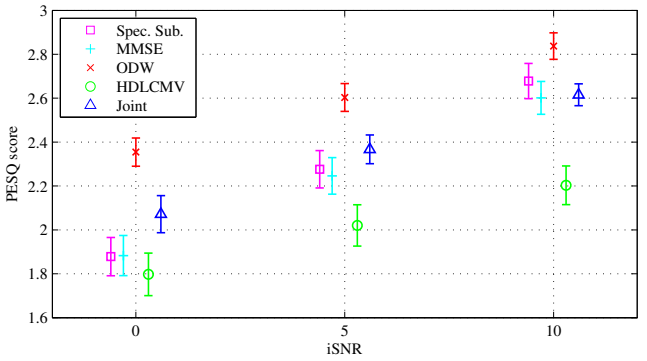
where $\mathbf{B} = \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}$ and $\mathbf{C} = \mathbf{Z}^H\mathbf{B}$. The harmonic distortion in (23) of the HDLCMV filter is always 0 due to its constraints.

4. JOINT ODW AND HDLCMV FILTERING

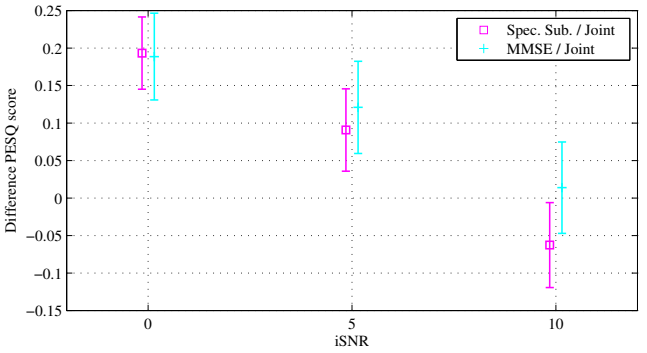
In this section, we propose to use the ODW and HDLCMV filters jointly for noise reduction in voiced speech segments. The joint use of the filters is relevant since they have complementary advantages and disadvantages. The ODW filter



(a)



(b)



(c)

Fig. 1. Average PESQ scores (a) for the joint filtering scheme as a function of M for an iSNR of 5 dB, and (b) for several enhancement methods as a function of the iSNR for $M = 110$ with 95% confidence intervals. In (c), the average differences in PESQ scores between the joint filtering scheme and the spectral subtraction and MMSE-based methods, respectively, are plotted with 95% confidence intervals.

is in practice reliant on the noise statistics. The noise signal is, however, not available directly in practice, so the noise statistics are relatively difficult to estimate. That is, nonstationary noise has a detrimental impact on the performance of the ODW filter. The HDLCMV filter, on the other hand, is driven by the pitch, and the observed signal statistics. It should therefore be more robust against nonstationary noise since the noise statistics are not needed directly in the filter design. The HDLCMV filter, however, assumes that the desired signal is quasi-periodic which is not exactly true for all parts of speech. As a result of that, distortion will be introduced by the HDLCMV filter due to model mismatch. Therefore, it should be beneficial to use the filters jointly. In the joint filtering scheme, the HDLCMV filter is used to obtain a rough estimate of the desired signal. This estimate is then subtracted from the observed signal to obtain an estimate of the noise. The estimated noise is used to find the noise statistics which, eventually, are applied in the design of the ODW filter. Finally, the ODW filter is utilized for estimating the desired signal.

The proposed joint filtering scheme was evaluated by measuring "Perceptual Evaluation of Speech Quality" (PESQ) scores [9]. The PESQ score is an objective measure that reflects the subjective quality of a speech signal, and the score can be measured relative to an original speech signal or not. That is, by evaluating the proposed scheme using PESQ scores, we evaluate the perceptual performance of the scheme. We compared the PESQ scores of the signals enhanced using the joint filtering scheme with those enhanced using the ODW filter only, the HDLCMV filter only, a spectral subtraction based method [10], and a method using MMSE spectral amplitudes [11]. In the design of the ODW filter, the noise signal is assumed available, so the performance of this method can be thought of as an upper bound on the performance of the proposed method. Followingly, we describe how the enhancement methods were set up for the evaluation. The statistics needed for the filter designs were replaced by the respective sample covariance matrices calculated from the past 400 samples. The filters in the joint filtering scheme were reguralized using [12]

$$\hat{\mathbf{R}}_{\text{reg}} = (1-\gamma)\hat{\mathbf{R}} + \gamma \text{Tr}\left\{\hat{\mathbf{R}}\right\} M^{-1}\mathbf{I}\,, \tag{28}$$

where $\text{Tr}\{\cdot\}$ is the trace operator and γ is the regularization factor. Regularization was necessary due to estimation error on the signal statistics and model mismatch. We chose $\gamma = 0.7$ which gave consistently good results in terms of PESQ scores. At each time instance, the model order was set to $L = \min\{[15, \lfloor \pi/\omega_0 \rfloor - 1\}$. The speech signals used for the evaluation contains both voiced and unvoiced parts, however, the HDLCMV filter in the proposed filtering scheme is suited for voiced speech enhancement only. Therefore, in the simulations, we updated the HDLCMV filter as follows; for voiced speech segments the HDLCMV filter was designed using (26) while, for unvoiced speech segments, it was updated as

$$\mathbf{h}(n) = (1-\lambda)\mathbf{0} + \lambda \mathbf{h}(n-1)\,, \tag{29}$$

when $\|\mathbf{h}(n-1)\|_2 > 0.1$ with $\lambda = 0.95$ and $\mathbf{0}$ is the zero vector. The spectral subtraction and MMSE based methods

are available in the VOICEBOX toolbox[1] for MATLAB in which they are implemented using noise power spectral density estimates calculated using optimal smoothing and minimum statistics [3]. We used the defaults settings in the toolbox for these enhancement methods.

We conducted a number of experiments where we used the joint filtering scheme for nonstationary noise reduction. For these experiments, we used two female and two male speech excerpts of length 4-6 seconds taken from the Keele database [13]. In this paper, we treat the pitch and the harmonic model order as known parameters to evaluate the maximum achievable performance of the proposed method. Therefore, we used the pitch information from the Keele database to design the HDLCMV filter. Moreover, we do not consider voiced/unvoiced speech detection in this paper. The pitch track from the Keele database contains zeros when the speech signal is unvoiced or no speech is present, so this information was used to circumvent the detection problem. We then generated observed signals by adding different noise types to the different speech excerpts; the added noise types were white Gaussian noise, car noise, babble noise, exhibition hall noise, and street noise. All noise sources except the white noise were taken from the AURORA database [14]. First, we enhanced the noisy signals at an iSNR of 5 dB at different filter lengths, and the PESQ scores were measured and average across the different excerpts. The resulting PESQ scores are shown in Fig. 1a. It can be seen that the perceptual performance is highest around $M = 110$. We then enhanced the noisy signals for different iSNRs when the filter length was $M = 110$. The average PESQ scores with 95 % confidence intervals are depicted in 1b. These results indicate that the proposed scheme outperforms the spectral subtraction and MMSE-based methods for iSNRs of 0 and 5 dB on average in terms of perceptual confidence. To investigate this further, we measured the average difference in PESQ scores between the proposed scheme and the two other methods; the average differences are shown in Fig. 1c. From these results, we can conclude that the proposed scheme outperforms the spectral subtraction and MMSE-based methods in terms of average PESQ scores with 95 % confidence for low SNRs.

5. CONCLUSIONS

In this paper, we proposed a joint filtering scheme for nonstationary noise reduction of quasi-periodic signals. The joint scheme consists of the ODW and HDLCMV filters. The ODW filter is driven only by the noise statistics and is therefore appropriate for enhancement of any desired signal. However, in practice the noise is not available directly, so the noise statistics are difficult to estimate. As a consequence of that, the performance of the ODW filter is deteriorated by nonstationary noise. The HDLCMV filter assumes that the desired signal is periodic and thereby has a harmonic structure. This is a good assumption for voiced parts of speech signals. Using this assumption, the HDLCMV filter is designed using the pitch and the model order of the desired harmonic signal, and the statistics of the observed signal, i.e., this filter is not dependent on the noise statistics. The HDLCMV filter is therefore more robust against nonstationary noise, but it will introduce some distortion in practice due to the periodicity assumption. The advantages and disadvantages of the ODW and HDLCMV filter are complementary, and we therefore proposed to use the filters jointly. In the joint scheme, the HDLCMV filter is used to estimate the noise statistics which are then used to design the ODW filter. The noise reduction is then performed by the ODW filter. We showed that the proposed joint filtering method outperforms existing speech enhancement methods in terms of average PESQ scores with 95 % confidence for relatively low iSNRs ($\leq$ 5 dB).

[1] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

6. REFERENCES

[1] P. Loizou, *Speech Enhancement: Theory and Practice*. CRC Press, 2007.

[2] D. Ealey, H. Kelleher, and D. Pearce, "Harmonic tunnelling: tracking non-stationary noises during speech," in *Proc. Eurospeech*, 2001.

[3] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[4] S. Ahmadi and A. S. Spanias, "Cepstrum-based pitch detection using a new statistical v/uv classification algorithm," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 3, pp. 333–338, may 1999.

[5] E. Fisher, J. Tabrikian, and S. Dubnov, "Generalized likelihood ratio test for voiced-unvoiced decision in noisy speech using the harmonic model," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 2, pp. 502–510, mar 2006.

[6] J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters – A Theoretical Study*, 1st ed., ser. SpringerBriefs in Electrical and Computer Engineering. Springer, 2011, no. VII.

[7] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.

[8] ——, "Optimal filter designs for separating and enhancing periodic signals," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 5969–5983, Dec. 2010.

[9] ITU-T, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," no. P.862, pp. 1–30, Feb. 2001.

[10] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 4, 1979, pp. 208–211.

[11] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[12] F. van der Heijden, R. P. W. Duin, D. de Ridder, and D. M. J. Tax, *Classification, Parameter Estimation and State Estimation - An Engineering Approach using MATLAB®*. John Wiley & Sons Ltd, 2004.

[13] F. Plante, G. F. Meyer, and W. A. Ainsworth, "A pitch extraction reference database," in *Proc. Eurospeech*, Sep 1995, pp. 837–840.

[14] D. Pearce and H. G. Hirsch, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. Int. Conf. Spoken Language Process.*, Oct 2000.